Using Machine Learning to Classify Data From the Physical World

Workshop by Lab 651
Annie Tomassoni
Mike Novi
Justin Grammens

Lab|651

Who is Lab 651

- Industry professionals helpling companies decrease costs, increase revenue, open new revenue streams with data or improve the quality of life using IoT technology
- Multi-disciplinary development team of Hardware, Mechanical, Firmware, Software with extensive experience in wireless communications
- Amazon AWS and Microsoft Azure Design partner
- Extensive expertise in cloud, mobile and web applications and data science
- Longtime sponsor and supporter of IoTFuse

Who are we?

- Justin Grammens
  - 20+ years architecting & engineering software
  - Cofounder of Lab 651
  - Cofounder of IoT Fuse
  - Owner of IoT Weekly News
  - Adjunct Professor at the University of Saint Thomas teaching graduate level course on IoT
- Mike Novi
  - 15+ years as architect and software engineer
  - Early hire at numerous startups: Massive Inc (acquired by MSFT), Statuspage (acquired by TEAM)
- Annie Tomassoni
  - Software Engineer at Lab 651

Lab|651

Reasoning for this Workshop

Lab|651

Reasoning for this Workshop

However…
Data is Not Useful without Insights

Lab|651

Scenario(s)

- You own a manufacturing company
    - Production line costs you $1000/hr when it's not operational
    - Biggest source of lost time and cost is machine failure
    - What if you could track machine behavior?
        - Predict movements that are indicators of failure
        - Order parts for maintenance during scheduled downtimes
        - Prevent wasting money doing maintenance when it's not needed
- You are a machine manufacturer
    - Currently you only receive revenue when the machine is purchased or need repair (low margin)
    - What if you could track machine behavior?
        - Offer a service to monitor and prevent downtime with your customer (high margin)
        - Better quality of service and experience by the customer using your product
- ATM reporting temperature example

Lab|651

What We'll Do

- For both scenarios we are really just sensing movement
- Phone as the proxy for the sensor
- Head as the proxy for the machine
- Classify data in the cloud as you move the phone
- Use Machine Learning to predict identify changes and classify changes
- Graph the data entering the system

Lab|651

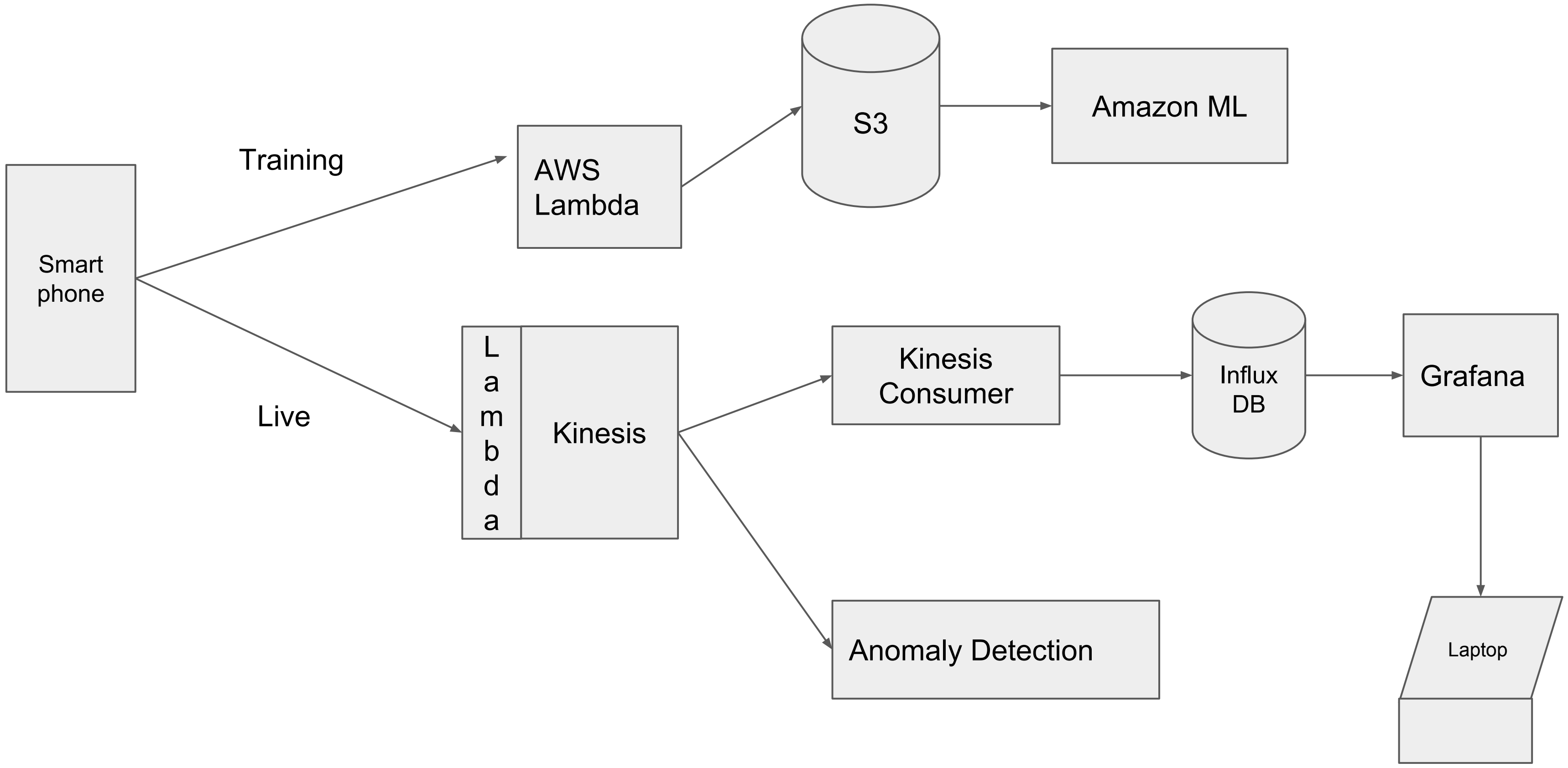
Technologies

- React Native
  - Framework for building native applications using javascript
- Amazon Web Services
  - Lambda
    - Let’s you run code without provisioning or running servers
  - Kinesis
    - Fully managed service to collect, process, and analyze real-time streaming data
  - S3
    - Object storage built to store and retrieve any amount of data from anywhere
  - Amazon ML
    - Service that makes it easy to use machine learning technology

Technologies

- InfluxDB
  - Open Source Platform built for metrics, events, and other time-based data
- Grafana
  - The open platform for beautiful analytics and monitoring
- Languages
  - Java
  - Javascript
  - Python

Lab 651: IoTFuse Workshop - 2018
Architecture Diagram



Cloud Provider Options

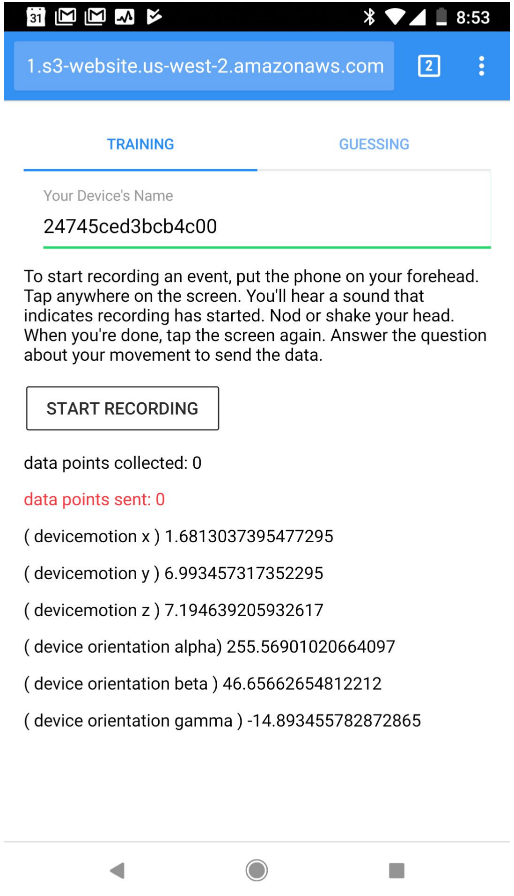
AWS	**Azure**	**Google**
Lambda	Azure Functions	Cloud Functions
EC2	Azure VM	Compute Engine
S3	Azure Storage	Cloud Storage
Kinesis	Stream Analytics	Cloud Dataflow
Amazon ML	Azure ML	Google ML Engine
QuickSight	PowerBI	Google Data Studio
DynamoDB	Azure Cosmos DB	Cloud BigTable

Cloud Benefits

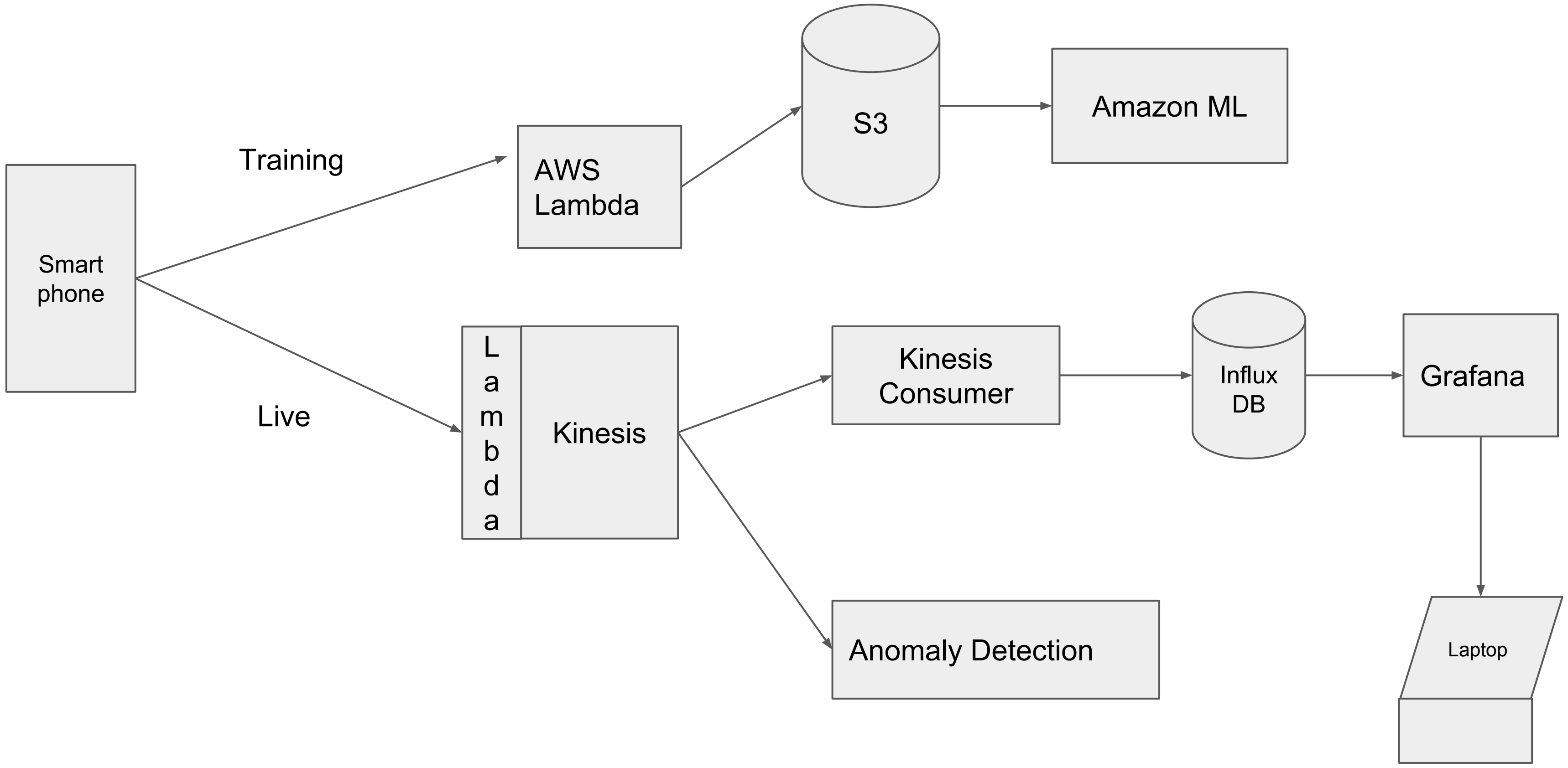
- IAAS (Infrastructure as a Service)
  - Burstable and will auto scale when needed
  - Serverless (only pay for what you use)
- Plug microservices together
  - Very little code actually needs to be written
  - Use your language of choice

Lab|651

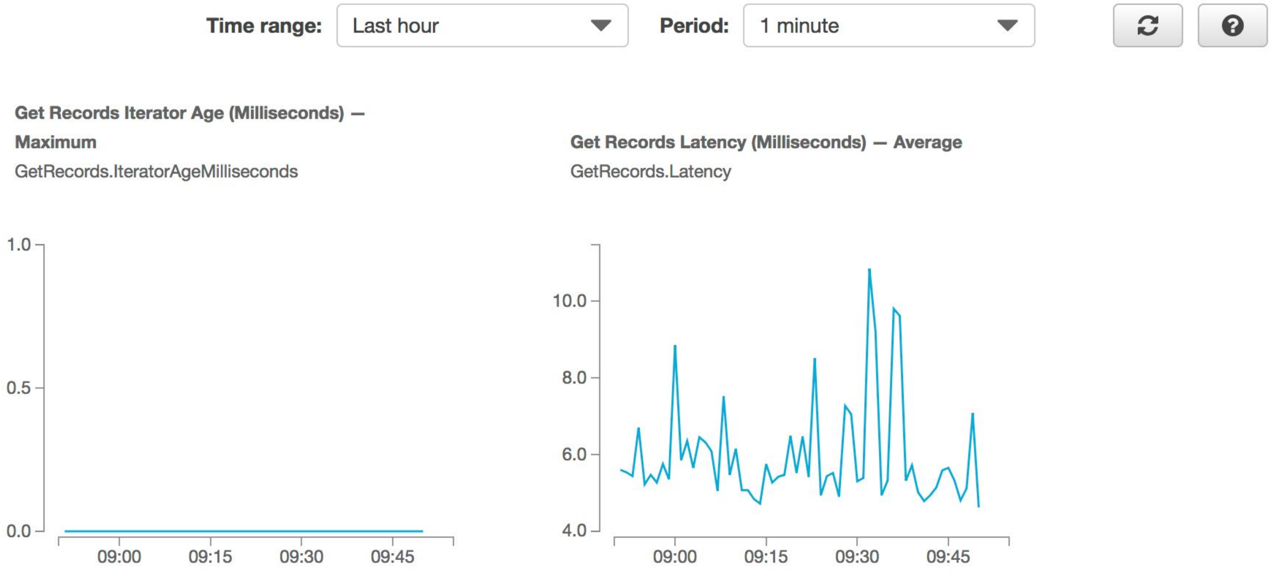
Mobile Application



Lab 651: IoTFuse Workshop - 2018
Architecture Diagram



Lamda & Kinesis



Show our AWS Console for Kinesis and Lambda

Influx and Grafana

Example JSON


```
{
    "acceleration": {
        "x": -2.7006595134735107,
        "y": 4.098873138427734,
        "z": 8.5903959274292,
        "timestamp": 1525049160042
    },
    "gyroscope": {
        "x": 65.03981000044783,
        "y": 23.49974914686357 5,
        "z": 15.860402996801678,
        "timestamp": 1525049160035
    },
    "device": "3d0744b3abe85600000000000000000"
}
```


Lab|651

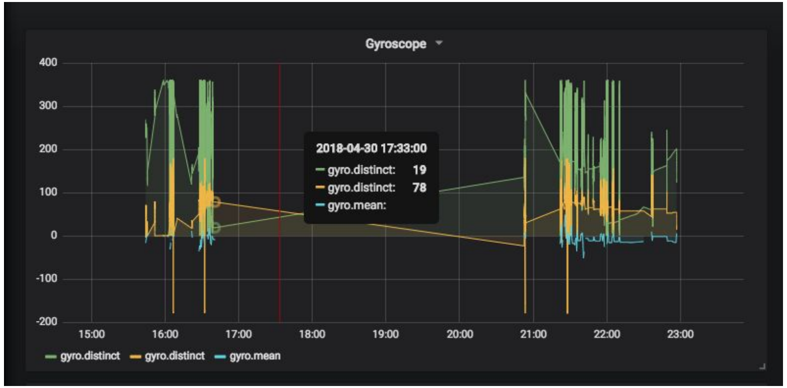
Influx

- Influx
  - Login to EC2 instance
  - Exec `infux`
  - Exec `Show databases`
  - Exec `Use iotfuse2018`
  - Exec `show measurements`
  - Exec `show field keys`
  - Exec `select * from acc limit 10` and `select * from gyro limit 10`
- Influx Query API
  - https://docs.influxdata.com/influxdb/v1.5/guides/querying_data/

Lab|651

Grafana

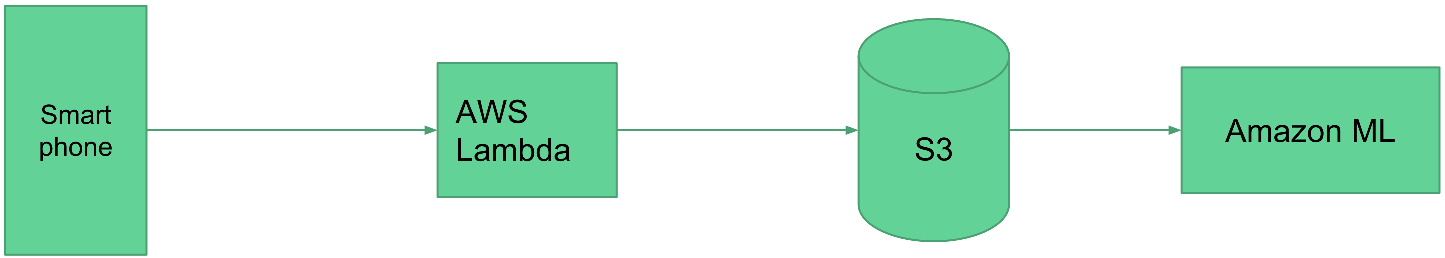
- sudo apt-get install grafana
- sudo service grafana-server start
- By default service runs on port 3000
  - Reads from a variety of time series databases
  - Allows for refresh as quick as 1 second



Lab|651

Building the Training Framework

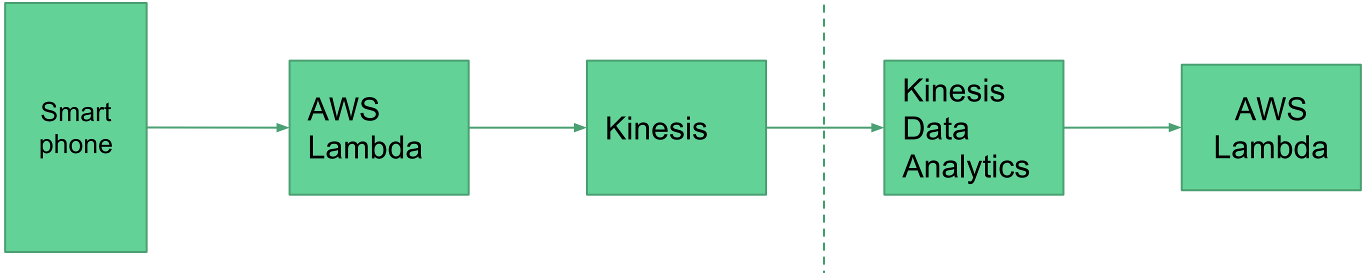
- https://github.com/lab651/Workshop-IoTFuse2018
- Using Amazon ML with supervised learning
    a. Collect Training Data (Data with Results)
    b. Store Data as Necessary
    c. Begin Collecting Training Data
    d. Prepare Data to Create Machine Learning Model
    e. Create Machine Learning Model / Data Source
    f. (optional) Setup the classification endpoint



Lab|651

Establishing the Live Data Pipeline

- Receive Constant Stream of Data
- Statistical Windows
- Classify the Motion (Yes/No/Unknown)



Options for Improvement

- Beyond simple Averages (FFTs)
- Anomaly Detection
  - Reduce cost / filter for changes]
    - Only run classification when changes occur
  - RANDOM_CUT_FOREST (available in Kinesis Data Analytics)
    - Works with ALL numeric data (including timestamps)
    - Often better at detecting sudden change
- Event Memory
  - Kinesis is noisy / “at least once” delivery
  - Mechanisms to filter out repeated data points
- More robust time series analysis (InfluxDB and tools)

Lab|651

Did We Get This Backwards?

- We don't always know what failures look like
- Unsupervised Learning
  - Anomaly detection
  - Cluster anomalies
  - Classify clusters

Lab|651